A robust supervised method for estimating soil moisture content from spectral reflectance

Bikram Koirala, *Member IEEE*, Zohreh Zahiri, Paul Scheunders, *Senior Member IEEE*

***Abstract*—Due to the complex interaction of light with moist soils, the soil moisture content (SMC) is hard to estimate from the soils spectral reflectance. Spectral variability, caused by variations in viewing and illumination angle and between-sensor variability further complicates the estimation. In this work, we developed a supervised methodology to accurately estimate SMC from spectral reflectance. The method determines a proxy for the SMC of a moist soil, making use of the reflectance spectra of an air-dried and saturated soil sample. The proxy is made invariant to illumination and viewing angle, and sensor type. In the next step, the proxy is normalized with respect to the ground truth SMC of the saturated soil to make the technique less dependent on the soil type. The normalized proxy can be directly used as an estimate of SMC. Alternatively, the nonlinear relationship between the normalized proxy and the actual SMC can be learned by supervised regression. Experiments are conducted on real moist soil data. In particular, we developed data sets of moist minerals, acquired by two different sensors, an Agrispec spectrometer, and an Imec snapscan shortwave infrared (SWIR) hyperspectral camera, under strictly controlled experimental settings. The proposed methodology is also validated on the available real moist soil data from the literature. Compared to state-of-the-art methods, the proposed method accurately estimates the soil moisture content.**

***Index Terms*—Remote sensing, Hyperspectral, Soil moisture content, Soil, Machine learning regression**

I. Introduction

Soil moisture is the main source of water for agriculture and natural vegetation ([1]). Soil moisture relates to the amount of water contained in the unsaturated soil zone ([2]). In [3], three main types of soil water are described: (a) hydration (absorbed) water incorporated into the lattice of the minerals; (b) hygroscopic (adsorbed) water bound to soil particles by the surface electrical charges of the minerals; and (c) free water covering the minerals and occupying the pores. Soil water plays a key role in crop production by serving as a solvent for nutrients such as sodium, potassium, carbon, nitrogen ([1]). It also plays a very important role in the exchange of mass and energy between the Earth's surface and the atmosphere ([2], [4], [5], [6], [7], [8]). Timely and accurate assessment of soil moisture content is critical in agriculture, hydrology, horticulture, geotechnical, and other environmental fields ([3], [4], [5], [9], [10]).

In practice, soil moisture content (SMC) is defined as the ratio of water to pure (dry) soil, either expressed in mass (gravimetric SMC) or volume (volumetric SMC) ([2], [4]). In the remaining of this manuscript, we will refer to gravimetric SMC as SMC, unless specified as volumetric SMC. The most accurate but intensive way of estimating soil moisture is the gravimetric method ([1], [11]), in which a fresh soil core sample is weighted and oven-dried until no further loss of mass is observed. The weight difference between the wet and dried sample is an estimate of the soil moisture. The most commonly used ground-based techniques (time-domain reflectometry, frequency domain sensors, etc.) estimate volumetric SMC ([2], [3], [9], [12], [13], [14]). Although these methods can be adapted to monitor temporal variations of volumetric SMC locally (in a small area), they cannot be applied to map spatial variations ([12]). To estimate volumetric SMC globally, a large amount of research has primarily been focused on the microwave part of the spectrum ([6], [7], [10]), because moisture strongly affects soil dielectric properties and longer wavelengths penetrate relatively deep into the ground ([10]). The major advantage of the microwave part of the spectrum is that vegetation on the surface does not influence the longer wavelength microwave radiation ([4]). Even though a large number of algorithms has been developed to exploit microwave radiation for volumetric SMC estimation at a large scale, the data products typically have coarse spatial resolution, even from airborne systems ([12]) and might be not useful for monitoring e.g., small catchment areas ([14]). Moreover, information is required regarding the soil surface roughness and the soil dielectric constant.

It is well-known that soil moisture largely influences the reflection from soil surfaces in the VNIR (400–1100 nm) and SWIR (1100–2500 nm) regions of the spectrum ([8], [15]). The main advantage of using this region of the spectrum is that the solar radiation works as a natural illumination source. Despite the limited light penetration depth (a few micrometers to a few millimeters), the SMC of the topmost layer of the soil can be determined at a high spatial resolution using hyperspectral imaging ([16]). In earlier work, the impact of soil moisture on reflectance was mainly studied qualitatively. Anders Ångström ([17]) experimentally observed a decrease in the reflectance when wetting soils artificially. He explained this by internal reflections of the reflected radiation in the water film covering the soil particles. Several authors reported the same ([18], [19]). These studies were mostly performed on unsaturated soils ([7], [20]). When the soil moisture content is over-saturated, the reflectance of a soil can be higher than that of the saturated soil, due to specular reflection.

More recently, a large amount of quantitative research was conducted to estimate SMC from reflectance measurements and results were promising for different types of soils ([10]).

Bikram Koirala, Zohreh Zahiri, and Paul Scheunders are with the Imec-VisionLab, Department of Physics, University of Antwerp, Belgium e-mail: {bikram.koirala, zohreh.zahiri, paul.scheunders}@uantwerpen.be

The methodologies developed in these works can be grouped into empirical methods and methods based on physical modeling. Empirical methods include spectral indices ([10], [21], [22], [23]), statistical relationships ([24], [25]), exponential functions ([7], [8], [26]), wavelet analysis ([27]), and multivariate analysis ([28]). Most of these methods utilize a limited number of wavebands for estimating SMC. For example, in [10], the normalized soil moisture index (NSMI) was introduced to estimate SMC. Most of these researches concluded that SWIR bands are better suited for estimating SMC, due to the strong absorption bands of molecular water around 1400 nm and 1900 nm ([6], [8]), while in the VNIR spectral region, the reflectance of the soil reaches its minimum value at much lower SMC ([8]). These methods were mostly validated on data acquired in a controlled laboratory environment. In field conditions, the use of the entire wavelength range is required since the strong absorption bands of molecular water around 1400 nm and 1900 nm are masked by atmospheric absorption ([6], [8]).

Most of the physical modeling techniques ([3], [29], [30]) utilize the entire VNIR and SWIR wavelength region (400-2500 nm). The main goal of these models is to physically describe the interaction of light with moist soils. A major advantage of the physical modeling techniques is that they can simulate the reflectance spectrum of the moist soil. In [29], the Bach model was introduced, which describes the reflectance of wet soil as a function of the soil reflectance in its dry form and a parameter L that denotes the active thickness of the water layer. Inversion leads to an estimate of L as a proxy for the SMC. A supervised regression model uses a training dataset with known SMC to learn the relationship between L and the SMC. In [3], MARMIT, an improved multi-layer Bach model was proposed.

An exception to the physical modeling techniques is the Sadeghi model ([16]), in that it neither simulates the reflectance spectrum of the moist soil nor utilizes the entire wavelength range. In [16], it was recommended to use the reflectance value at the wavelength 2210 nm. The model estimates the SMC from the soil reflectance by making use of the reflectance of a dry soil sample and a saturated soil sample.

Both empirical and physical modeling techniques have their disadvantages. Two major challenges remain for an accurate estimation of SMC:

- Variations in illumination and viewing angle and sensor type cause a huge variation in the reflectance spectra of moist soil samples. Some effects cause a mere scaling of the reflectance spectra (e.g., illumination and viewing angle), others cause a multiplication with a wavelength-dependent function (cross-sensor situations, e.g., difference in sensor type, white calibration). Most of the physical models are not invariant to these effects. Although spectral indices-based methods are invariant to scaling effect caused by variations in illumination and viewing angle, they are not invariant to cross-sensor situations.
- The spectral reflectance of a moist soil highly depends on the soil grain size (texture) and grain size distribution. These determine the distribution of the water in the soil. For two soils with identical SMC, the incident light travels longer distances in the water before reflecting out from the sample (i.e., the optically active thickness of the water layer) in the soil containing the largest pore size. This effect is visible e.g., as differences in the depth of the absorption bands of molecular water around 1400 nm and 1900 nm (see e.g. Fig. 3). The existing methods do not properly consider the distribution of the water in the soil, making the results highly dependent on the soil type. Spectral indices and the Sadeghi model use a limited number of bands, making it hard to properly capture the distribution of the water in the soil. The Bach and MARMIT model employ the optically active thickness of the water layer as a proxy for the SMC, which is soil-dependent. As a consequence, when trained on one soil type, the models are not transferable to other soil types. Although the distribution of the water in the soil is partially characterized by the spectral reflectance of the saturated soil, it can vary significantly during drying because of the capillary and adsorptive forces. One observable effect in lab measurements is a random scaling of the spectral reflectance of the moist soils (especially non-hydrophobic soils) due to changes in the thickness of the sample.

In this work, we will focus on these two challenges. Our method consists of three steps. In the first step, a proxy for SMC is generated by describing the moist reflectance relative to the reflectance of both a dry and a saturated soil sample. The soil textural information and the distribution of the water in the soil are partly implicitly embedded in the spectral reflectance of the saturated soil. The proxy for SMC is made invariant to random scaling effects, including variations in illumination and viewing angle, as well as sensor type variations. In the next step, the proxy is normalized with respect to the ground truth SMC of the saturated soil to make the technique less dependent on the soil type. Indeed in the methodology section (see Figure 7), we will demonstrate this. In the final step, to estimate SMC, we will propose two different approaches. As the relation between the normalized proxy and SMC will be shown to be nearly linear, in the first approach, the normalized proxy will be directly used as an estimate of SMC. To improve accuracy, in the second approach, a supervised regression algorithm is applied to learn the actual relationship between the normalized proxy and the SMC of the moist soil. Although in the experimental part, we will demonstrate that the regression parameters obtained by training the algorithm on one soil type can be utilized to estimate soil moisture content of any other soil type without further calibrating for new soil types, the second approach is more suitable for soil-specific calibration. The proposed methodology will be validated experimentally on real moist soil datasets from the literature ([3]), and a self-crafted hyperspectral dataset of moist minerals.

The remaining of this article is organized as follows. In Section II, we describe our dataset and the datasets from the literature on which our methodology is validated. Section III is devoted to prior work on some of the physical models that

TABLE I: Density and grain size (D50) of the minerals used in the Imec-Visionlab_2021 dataset.

Mineral	Density (g/cm^3)	Grain size (μm)
FeO(OH)	4.25	44
Fe_2O_3	5.26	0.8
Lime	2.7	-
SiO_2	2.64	23
WS	2.64	-

estimate SMC of moist soils. In Section IV, the proposed methodology is elaborated. In Section V, we describe the experiments and the results, followed by a discussion in Section VI. Section VII concludes this work.

II. Moist soil datasets

A. Imec-Visionlab_2021

The dataset contains five dry minerals in powder form (see Table I), typically found in natural soil on the Earth: Goethite (FeO(OH)), Iron oxide (Fe_2O_3), Lime, Silicon dioxide (SiO_2), and White Sand (WS). Lime (calcium/magnesium-containing powders) is often added to the soil to reduce the acidity of the soil.

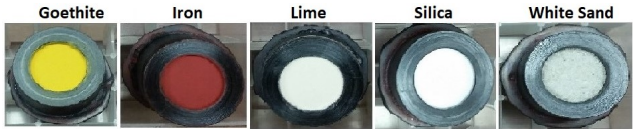


Fig. 1: Minerals of the Imec-Visionlab_2021 dataset

Each sample was put inside a round black sample holder with an interior diameter of 20 mm, a height of 5 mm, and an edge thickness of approximately 3 mm (see Figure 1). The sample holder was filled and compacted and smoothened using a stamp compactor. Before wetting these samples, the reflectance spectra were measured by an AgriSpec spectrometer (manufactured by ASD (Analytical Spectral Devices)) and a snapscan hyperspectral SWIR camera (manufactured by Imec). The data from the spectrometer have 1500 spectral bands, ranging from 1000 nm to 2500 nm with a step size of 1 nm. The spectral range of the camera is 1100-1670 nm with a spectral resolution of approximately 5 nm, resulting in a total of 113 spectral bands. In contrast to push broom systems, in which either the camera or the sample should move, the sensor moves inside the snapscan camera, allowing it to acquire a still full image frame. The original frame size of the raw images was 100 $\times$ 100 pixels. To provide data with uniform illumination and remove unrelated objects (edge of the sample holders) the images were clipped to 30 $\times$ 30 pixels. Since all samples were homogeneous, no spatial variation between the spectra was observed, and the spectra of all pixels were averaged over the entire clipped image. Figure 2 shows the spectra of the air-dried minerals acquired by both the spectrometer and the camera. A substantial difference between the acquired spectra of the different sensors can be observed, due to external variability, including variations in illumination and distance of the samples to the sensor, causing global scaling effects and wavelength-dependent variations in the acquired spectral reflectance.

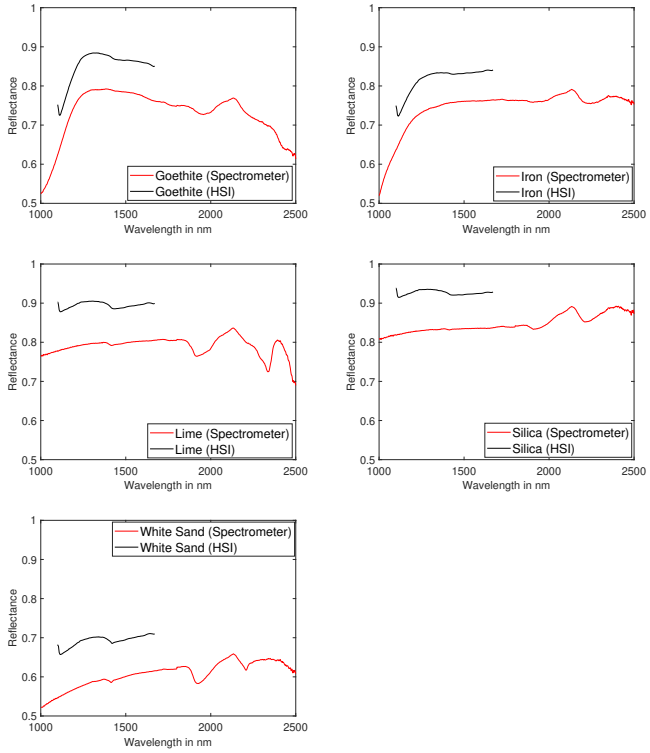


Fig. 2: Spectra of the air-dried minerals of the Imec-Visionlab_2021 dataset, acquired by the spectrometer (red line) and hyperspectral camera (black line).

Next, water was slowly poured down from the top, until saturation (when a layer of water starts to appear at the top level). Water and minerals were mixed homogeneously with the help of a spatula. Approximately 5-10 mg of the mineral was found to stick on the spatula during mixing and was considered when calculating SMC. During drying, the spectra and the weight of the samples were regularly measured until the final weight of the sample returned to its initial value (after approx. 20 hrs). For each measurement, the acquisition of the reflectance spectra by both the spectrometer and the camera took approximately three minutes. The time interval between two different measurements was approximately 15 minutes. As the original sample was slightly over-saturated, the first few measurements were removed, leading to approximately 40-60 different moisture levels for each mineral. SMC (θ) was calculated by applying the following formula:

$$\theta = \frac{(m - m_0)}{m_0} \times 100 \quad (1)$$

where m is the measured mass of the moist soil sample and m_0 is the initial (air-dried) mass. The resolution of the balance (Sartorius) is 0.1 mg. Remark that soils with large porosity can have mass ratio's larger than 100 %. In Table II, we summarize this dataset.

Figure 3 shows the spectral reflectance of all measured samples of the five minerals and their associated SMCs.

TABLE II: Summary of the dataset. Here, N denotes the number of reflectance spectra that was acquired during the course of the drying process.

Mineral	N	SMC (g/g $\times$ 100)	Drying protocol
FeO(OH)	52	0-101 %	Air-dried
Fe_2O_3	53	0-47 %	Air-dried
Lime	43	0-48 %	Air-dried
SiO_2	62	0-93 %	Air-dried
WS	47	0-44 %	Air-dried

Saturated soils are represented by dark red curves, while air-dried minerals are represented by black curves. To illustrate the change in the spectral reflectance during the air-drying process, the reflectance spectra of moist minerals at all intermediate SMCs are shown.

B. Liu_2002

A large number of soil samples (92 samples) was collected at several places in the world ([7]). After air-drying these samples, they were passed through a 2 mm sieve. Each soil sample was put into a box (radius 5 cm) forming a 1.5 cm thick layer, considered as optically infinitely thick. To make the samples fully saturated, water was slowly poured down from the side of the box. After free water disappeared from the soil surface (about 24h after the pouring), the drying process started, during which the spectral reflectance of each sample was measured frequently. An ASD Pro FR Portable Spectroradiometer (350–2500 nm) was used to acquire the reflectance spectra, with a spectral resolution of 1.4 nm in the shorter wavelengths and up to 2 nm for the longer ones. For each soil sample, four moisture levels, in the range 0-81 (g/g $\times$ 100) were measured. For each moisture level, spectra were acquired using 28 different acquisition configurations, by varying the illumination angle from 0^0 to 60^0 and the viewing angle from -45^0 to 45^0.

C. Lobell_2002

4 different soils: Argic Aridisol, Xeric Andisol, Ustic Mollisol, and Aridic Entisol were collected from different parts of the USA ([8]). These soils span a large range of bulk densities, porosities, and organic C content. Each soil was passed through a 2 mm sieve and then oven-dried at 70^0C for two weeks. A 1 mm thick sample of each soil was placed in a 5.0 cm wide round aluminum tin. An Analytical Spectral Devices spectrometer (350–2500 nm) was used to measure the spectral reflectance of these samples. After capturing spectra from the dry samples, the samples were humidified until saturation. During drying, a number of spectral measurements was performed until the soil mass returned to its initial value. On each sample, between 9 and 15 moisture levels were measured in the range 0-109 (g/g $\times$ 100).

D. Lesaignoux_2008

32 soil samples were collected from eight locations in France, representing large soil property variability in visual coloration and texture ([10], [24]). Each soil sample was cleared from roots and gravel and put in a Petri dish with a diameter of 6 cm and thickness of 1 cm. The samples were humidified until saturation and then dried progressively in a laboratory oven (333.15 K). After each drying step of 30 minutes, the reflectance spectrum of each soil sample was measured by using an ASD FieldSpec Pro FR spectroradiometer (400-2500 nm) and weighted to obtain the SMC. For each soil sample, six moisture levels in the range 0-87 (g/g $\times$ 100) were measured.

E. Philpot_2014

3 different soils: Ithaca soil, Quartz sand, and Masonry sand were collected to represent three texture groups ([12]). Ithaca soil contains much clay and organic matter. Quartz sand consisted of 90 % white translucent silica while Masonry sand is generally light brown with black particles. In comparison to Ithaca soil, both Quartz sand and Masonry are moderately homogeneous. The sample holder (1.2 cm tall black plastic cylinder with a 5.1 cm inner diameter and a sieve bottom) was filled with dry soil and the surface was leveled with a metal straight edge. Instead of pouring water from the top surface to make the soil sample saturated, it was placed on a saturated sponge and the sieve bottom allowed water to be drawn up into the soil sample via capillary action. The weight of the sample holder was automatically measured during the entire experiment by putting it on a scale (Ohaus SP200) with an accuracy of 0.01 g. The reflectance spectra were measured every 5 min intervals by using an ASD FieldSpec Pro FR spectroradiometer (350–2500 nm) until the soil mass returned to its initial value. The number of moisture levels for each sample varied between 97 and 205 in the range 0-50 (g/g $\times$ 100).

F. Bablet_2016

17 soil samples were collected from Plaine de Versailles and Tunisia and were passed through a 2 mm sieve ([3]). The samples were humidified until saturation and then dried progressively in an oven to obtain successive SMC values. An ASD FieldSpec Pro FR spectroradiometer (400–2400 nm) was used to measure the spectral reflectance of these samples. The number of moisture levels for each sample varied between 6 and 8 in the range 0-79 (g/g $\times$ 100).

III. Relevant Prior Work

A. The Sadeghi model (SM)

The Sadeghi model ([16]) was derived from the Kubelka–Munk (KM) theory ([31]). The KM model is a radiative transfer model that describes the interaction of light with an infinitely thick medium, containing uniformly distributed absorbing and scattering particles. It relates R, the reflectance of the mixture with r, the ratio between the light absorption coefficient (m^{-1}) and the light scattering coefficient (m^{-1}) by the following equation:

$$R = 1 + r - \sqrt{r^2 + 2r} \tag{2}$$

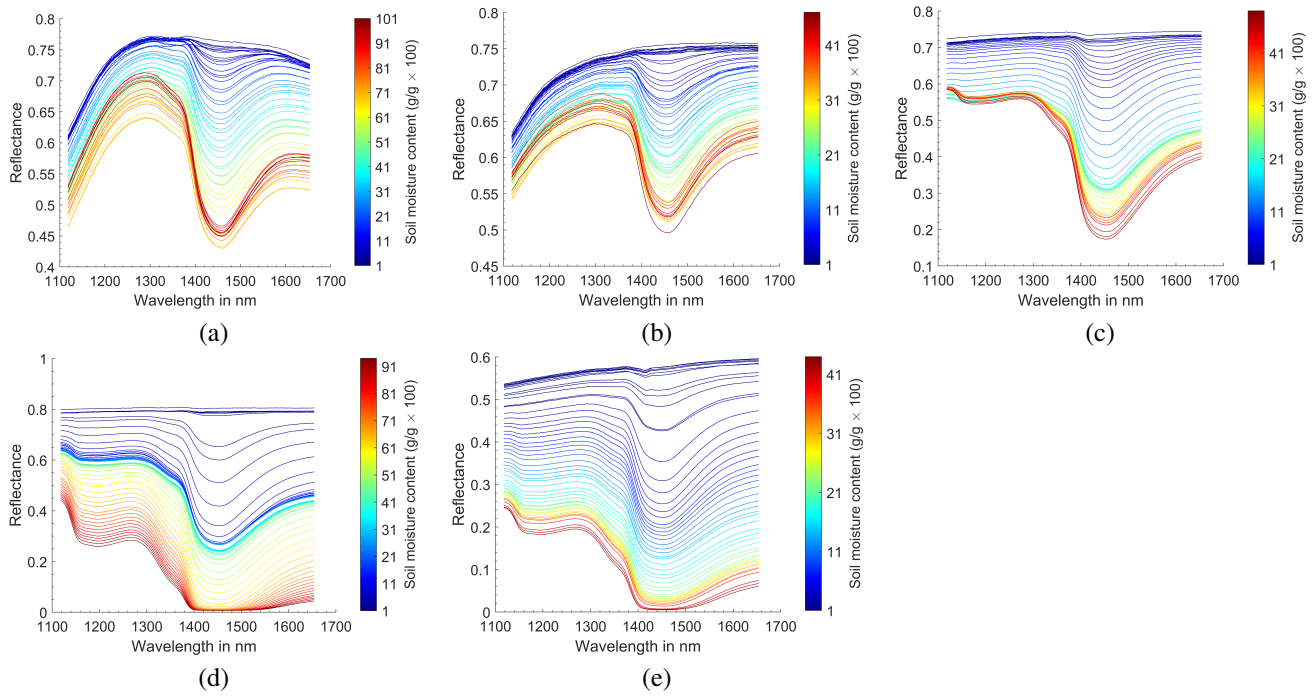


Fig. 3: Measured reflectance spectra (hyperspectral camera) and corresponding SMCs (see colorbar): (a) Goethite, (b) Iron, (c) Lime, (d) Silica, and (e) White sand.

Inversion leads to: $r = \frac{(1-R)^2}{2R}$. In [16], the SMC is obtained as:

$$\frac{\theta}{\theta_s} = \frac{r - r_d}{r_s - r_d} \tag{3}$$

where θ and θ_s are SMC of the moist and the saturated soil respectively, while r, r_d and r_s are the absorption/scattering ratio's of the wet soil, the dry soil and the saturated soil respectively. The main disadvantage of the Sadeghi model is that it is applicable only for a single (in principle any) wavelength. In [16], wavelength 2210 nm was recommended.

B. The Bach model

The Bach model ([29]) was developed by accounting for two processes: 1) the internal reflection of the reflected radiance in a water layer covering the soil, which reduces the soil reflectance:

$$R = \frac{R_d}{n^2 \times (1 - R_d) + R_d} \tag{4}$$

where R and R_d are the wet and dry soil reflectance, and n is the refractive index of water. 2) the absorption of the water that is bound to the soil particles, which can be modelled with Lambert's law:

$$R_B = R \times \exp(-2\alpha \times L) \tag{5}$$

where R_B is the soils total reflectance, α is the empirical absorption coefficient of water, and L is the active thickness of the water layer. By assuming that the soil surface is a patchwork of wet and dry areas, an efficiency term $\epsilon \in [0, 1]$ was introduced [32]:

$$R_{modB} = \epsilon \times R_B + (1 - \epsilon) \times R_d \tag{6}$$

L is a proxy for the SMC. To establish a relation between L and the SMC, a linear regression is performed on training spectra with known SMC.

C. MARMIT

The MARMIT model ([3]) is a multilayer radiative transfer model to estimate surface SMC. This model describes a wet soil as a dry soil, covered with a thin film of water (see Figure 4). This model derives the total reflectance of the water/soil

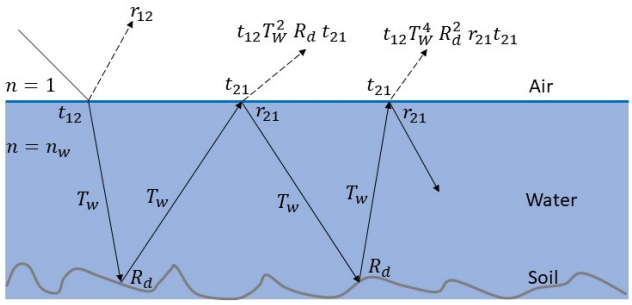


Fig. 4: A thin water layer on top of a rough soil surface. r_{12} and r_{21} are the reflection coefficients at the air-water interface and water-air interface respectively, t_{12} and t_{21} are the transmission coefficients at the air-water and water-air interface respectively, T_w is the transmittance of the water layer, R_d the reflectance of the dry soil, and n_w is the refractive index of water ([3])

.

system by summing the amplitudes of successive reflections

and refractions at the top of the water layer:

$$\begin{aligned} R_M &= r_{12} + t_{12}t_{21}R_dT_w^2(1 + r_{21}R_dT_w^2 + r_{21}^2R_d^2T_w^4 + \dots) \\ &= r_{12} + \frac{t_{12}t_{21}R_dT_w^2}{1 - r_{21}R_dT_w^2} \end{aligned} \tag{7}$$

where r_{12} and r_{21} are the reflection coefficients at the air-water and water-air interface, t_{12} and t_{21} are the transmission coefficients at the air-water and water-air interface. An expression of the reflection and transmission coefficients as functions of n, the refractive index of water is provided in [3]. $T_w = \exp(-\alpha L)$ is the transmittance of the water layer, with α the absorption coefficient of pure liquid water, provided by [33]. By assuming r_{12} to be zero, and introducing an efficiency term $\epsilon \in [0, 1]$, the following equation was derived:

$$R_{modM} = \epsilon \times \frac{t_{12}t_{21}R_dT_w^2}{1 - r_{21}R_dT_w^2} + (1 - \epsilon) \times R_d \tag{8}$$

The mean water thickness $L \times \epsilon$ is used as a proxy for the SMC. Based on training data with known SMC, a statistical relationship between $L \times \epsilon$ and SMC is established, using a sigmoid function:

$$SMC = \frac{K}{1 + a\exp(-\psi(L \times \epsilon))} \tag{9}$$

where K denotes the maximum value of the curve, a a translation factor, and ψ denotes the steepness of the curve.

Although all these models were developed based on physical reasoning and validated on datasets, acquired in a controlled environment, they may produce large errors when the data is impaired due to variations in illumination and viewing angles. These effects mostly cause scaling effects on the reflectance spectra. Even though these effects can be suppressed by normalizing the data for example by applying the standard normal variate (SNV), the physical models are not invariant to such transformations.

IV. A supervised method for estimating SMC from spectral reflectance

As we already mentioned in Section I, the spectral reflectance of moist soil is highly affected by the distribution of the water in the soil. This can be observed e.g., in absorption bands of molecular water around 1400 nm and 1900 nm. For soils with large grain sizes (e.g., white sand), for which the pore size is larger than the wavelength of light, the water absorption band around 1200 nm becomes also visible (Figure 3(e)). For soils with smaller grain size (e.g., iron), the water absorption bands are less deep (Figure 3(b)).

For a given soil sample, an increase in the SMC amounts to an increase in the depth of the water absorption bands, the deepest valleys are obtained when the soil is saturated. Thus, to properly capture the distribution of the water in a moist soil, one should describe its water absorption behaviour relative to that of its saturated version.

A. Spectral mixture analysis

In this work, we develop a supervised method to estimate the SMC of moist soil that utilizes the spectral reflectance of the air-dried and saturated soil and their ground truth SMC. The main idea is to regard a moist soil as a binary mixture of an air-dried and a saturated sample of the same soil. In this way, the problem can be approached similarly as with spectral unmixing of binary mineral mixtures, which is a highly nonlinear problem ([34]).

Let us assume that the two endmembers, i.e., the spectrum of the air-dried soil and saturated soil (with known SMC) are available. As the data manifold, sampled by a number of moist soils with varying SMC is a (nonlinear) curve in spectral space between the two endmembers, the relative arc length between a moist soil and the two endmembers can be regarded as a proxy for its SMC. This relative arc length can be calculated by approximating the curve by a piece-wise linear curve and by summing up the Euclidean distances between neighboring samples (see [34] for detailed information). The more moist soil samples are available, the better the approximation. In practical situations, only one moist soil (and the two endmembers) may be available, so that the approximation for the arc length leads to errors.

All spectra can be projected onto the unit sphere ($R = 1$) by normalizing the spectra, i.e., by dividing each spectrum by its length ($\mathbf{a} \rightarrow \frac{\mathbf{a}}{||\mathbf{a}||}$). On the unit sphere, the arc length between any two spectra is simply given by the angle between them, and can be computed by just calculating the arc cosine of their dot product. However, it is not guaranteed that, after projection, all data points lie on the arc connecting the two endmembers. In figure 5, we demonstrate a scenario where the spectrum of the moist soil $\mathbf{y}$ does not lie on the arc connecting the two endmembers $\mathbf{R_s}$ and $\mathbf{R_d}$ (red curve). To estimate the true arc lengths, the data point $\mathbf{y}$ has to be projected on the red curve resulting a new data point $\mathbf{y}'$. By utilizing the law of cosines (see equation 10), we can estimate the true arc lengths without projecting the data points on the red curve:

$$\begin{aligned} \cos\,(c) &= \cos\,(d)\,\cos\,(b_1) \\ \cos\,(c') &= \cos\,(d)\,\cos\,(b_2) \end{aligned} \tag{10}$$

After some calculations, one obtains from Eq. 10:

$$\cos\,(b_1) = \frac{\sin(b_1 + b_2)}{\sqrt{\left[\left[\frac{\cos\,(c')}{\cos\,(c)} - \cos\,(b_1 + b_2)\right]^2 + \sin^2(b_1 + b_2)\right]}} \tag{11}$$

where $b_1 + b_2 = \arccos(\mathrm{R}_d^T\mathrm{R}_s)$.

The relative arc lengths are then obtained as:

$$\hat{\mathbf{a}} = \begin{bmatrix} \frac{b2}{b1+b2} \\ \frac{b1}{b1+b2} \end{bmatrix} \tag{12}$$

where $\hat{\mathbf{a}}$ is the vector containing the relative arc lengths of the sample between the saturated and the dry endmember respectively.

The projection onto the unit sphere not only allows a direct calculation of the relative arc length of a moist sample, it also automatically becomes invariant to any random scaling of the spectrum of the measured sample, caused e.g., by illumination conditions. To demonstrate this, Figure 6 (a)) shows the scat-

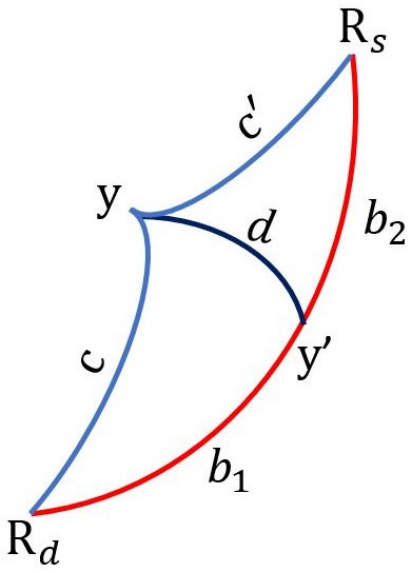
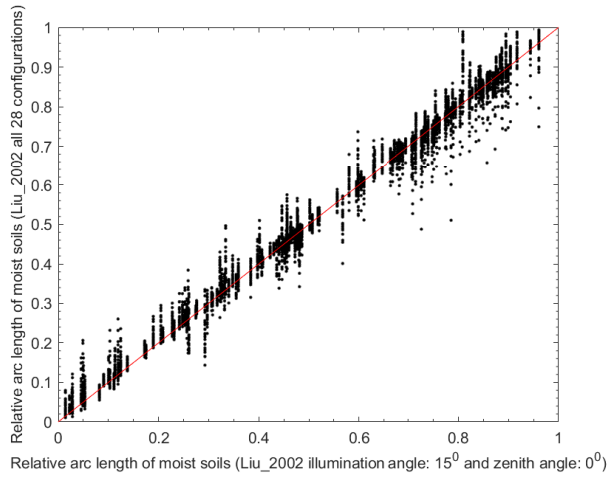


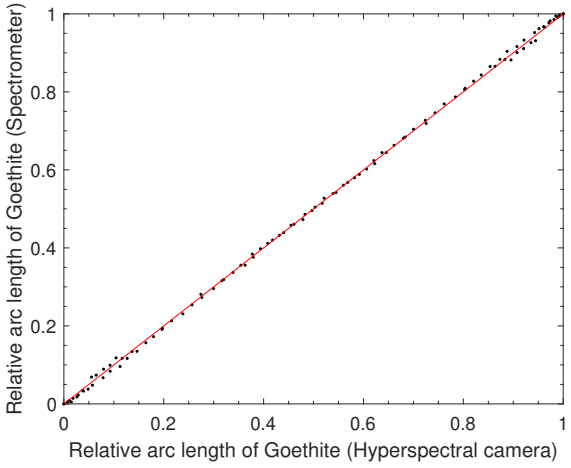
Fig. 5: Red curve: arc connecting the two endmembers; Blue curves: the arcs connecting the moist soil ($\mathbf{y}$) with the endmembers. c and c' denote the arc lengths between $\mathbf{y}$ and the endmembers $\mathbf{R}_d$ and $\mathbf{R}_S$, respectively. $\mathbf{y}'$ denotes the projection of $\mathbf{y}$ on the arc, and b_1 and b_2 denote the true arc lengths between $\mathbf{y}'$ and the endmembers.

terplot of the relative arc lengths of the soils of the Liu_2002 dataset for a particular configuration (illumination angle: 15^0 and zenith angle: 0^0) vs. all 28 configurations. Moreover, in all cases the endmembers are taken as the air-dry and saturated soil spectra at the illumination angle: 15^0 and zenith angle: 0^0. As long as the spectra of the moist soil and both the air-dry and saturated soil are measured with the same equipment (same sensor, same white calibration, ...), this representation is also invariant to cross-sensor situations. To demonstrate this, in figure 6 (b), the estimated relative arc lengths of moist Goethite, measured by the hyperspectral camera are plotted against the ones from the spectrometer measurements (Imec-Visionlab_2021 dataset).

A direct relation exists between the relative arc lengths and the SMC. This relation is expected to be similar for all soils that can hold the same amount of water in their saturated form. This amount however substantially differs between soils (see Table II). When normalizing the relative arc length with respect to the SMC of the saturated soil, by multiplying the relative arc length of the moist soil ($\hat{\mathbf{a}}$) by the SMC of the saturated soil (θ_s) ($\hat{\mathbf{a}} \rightarrow \hat{\mathbf{a}} \times \theta_{\mathbf{s}}$), a more general relation between the arc length and the SMC is obtained, that is similar for different soils. In Figure 7, this is demonstrated, by plotting the normalized relative arc lengths (NRAL) of moist soils vs. measured ground truth SMC of all soils involved in this study, i.e., from the five datasets from the literature and the five minerals from our own dataset. It can be observed that all points follow the same data manifold. The random scattering around the data manifold is due to complex soil-water-light interactions, which are not accounted for in this method. In general, we can claim that NRAL is a nearly



(a)



(b)

Fig. 6: (a) The estimated relative arc lengths of moist soils of the Liu_2002 dataset (Illumination angle: 15^0 and zenith angle: 0^0 vs. all 28 configurations); (b) The estimated relative arc lengths of moist Goethite (hyperspectral camera vs. spectrometer)

.

invariant measure of SMC. Among 153 soil samples studied in this work, silica deviates most (see section VI). From Figure 7, it is quite clear that the relationship between measured ground truth SMC and NRAL is slightly nonlinear. When assuming that the relationship is linear, NRAL can act as an estimate of SMC. In the experimental part, we will show the performance of the proposed method when NRAL is used as an estimate of SMC. However, to refine the estimation accuracy, the nonlinear relationship between NRAL and measured ground truth SMC should be learned.

B. Map normalized arclength to SMC

To estimate the SMC (θ) from NRAL accurately, a supervised approach is required to learn the mapping between both. For this, a set of n training samples with known θ: $\mathcal{D} = \{(\mathbf{y}_1, \theta_1), \ldots, (\mathbf{y}_n, \theta_n)\}$ should be available. Here, $\mathbf{y}_i$

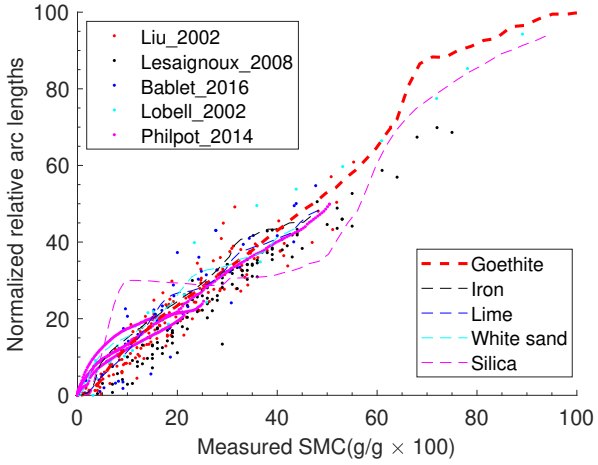


Fig. 7: The normalized relative arc lengths versus the ground truth SMC of all moist soils involved in this study.

indicates the reflectance spectrum of the i-th moist soil training sample.

1) Lookup table-based method: The simplest approach is to generate an ordered lookup table containing the normalized relative arc lengths and their corresponding SMC of all training samples. To estimate the SMC of a test sample, the normalized relative arc length is matched to the lookup table values, and eventually linearly interpolated in between two values. This approach assumes that the lookup table contains enough training data.

2) Supervised regression: The better approach is to apply a supervised regression model on the training data. In this work, we choose the Gaussian process as a regression model [35]. The estimated map of the normalized relative arc lengths of N test samples ($\hat{\mathbf{A}}_t = \{\hat{\mathbf{a}}_i\}_{i=1}^N$) to their SMC ($\mathbf{\Theta}_t = \{\hat{\theta}_i\}_{i=1}^N$) is given by:

$$\mathbf{\Theta}_t \quad = \quad \mathbf{\Theta}_{\mathcal{D}}(K(\hat{\mathbf{A}}_{\mathcal{D}}, \hat{\mathbf{A}}_{\mathcal{D}}) + \sigma_n^2\mathbf{I})^{-1}K(\hat{\mathbf{A}}_t, \hat{\mathbf{A}}_{\mathcal{D}})^T \quad (13)$$

Here, $\mathbf{\Theta}_{\mathcal{D}}$ is a row vector containing training SMC, $K(\hat{\mathbf{A}}_{\mathcal{D}}, \hat{\mathbf{A}}_t)$ is the matrix of kernel functions between the training normalized relative arc lengths ($\hat{\mathbf{A}}_{\mathcal{D}}$) and the test normalized relative arc lengths, and $K(\hat{\mathbf{A}}_t, \hat{\mathbf{A}}_t)$ is the matrix of kernel functions between the test normalized relative arc lengths. σ_n^2 is the noise variance of the normalized relative arc lengths in the training set. The kernel function is computed by the following equation:

$$k(\hat{\mathbf{a}}_i, \hat{\mathbf{a}}_j) = \sigma_f^2 \exp\left(-\sum_{b=1}^{p} \frac{\left(\hat{a}_i^b - \hat{a}_j^b\right)^2}{2l_b^2}\right) \quad (14)$$

where σ_f^2 is the variance of the input normalized relative arc length, $p = 2$ is the number of endmembers and l_b is a characteristic length-scale for each endmember. The hyperparameters of this kernel function are optimized by minimizing the negative log marginal likelihood of the training dataset (-log($p(\mathbf{\Theta}_{\mathcal{D}}^T|\hat{\mathbf{A}}_{\mathcal{D}}^T)$)).

Remark that the entire procedure of calculating the arc lengths, including the projection and normalisation generates a universal representation for the SMC, invariant to scaling effects, cross-sensor situations and partially accounting for the distribution of the water in the soil (see Figure 7). In the next section, We will demonstrate that the model can be trained with any (or multiple) soil types for which ground truth is available and is transferable to other soil types, irrespective of the sensor type or acquisition conditions.

In the remaining of this work, we will refer to this methodology as the supervised method for estimating SMC from spectral reflectance (SM_S).

V. Experiments and Results

The proposed methods NRAL and SM_S are validated and compared to two physical models and a method based on a spectral index:

- NSMI: The normalized soil moisture index [10]. This method relates the spectral reflectance to SMC by the following equation:

$$\text{SMC} = \left(\frac{R(1800\text{ nm}) - R(2119\text{ nm})}{R(1800\text{ nm}) + R(2119\text{ nm})} - a\right)/b \quad (15)$$

  where the values $a = 0.032$ and $b = 0.00897$ were obtained by a regression model, trained on a large number of soils. This index is scale invariant. However, it only uses information from two specific wavelengths, and it does not account for the distribution of the water in the soil to make the technique less dependent on the soil type.
- SM: Although the linear Sadeghi model is applicable to any of the SWIR wavelengths, it was recommended to use the reflectance value at the wavelength 2210 nm, as it is better correlated with the soil moisture content in comparison to other bands (see [16], Fig. 11). This method is not invariant to scaling, but it partially accounts for the distribution of the water in the soil, as it makes use of the spectral information of a saturated soil. Its main limitation is that it only uses information from one particular wavelength.
- MARMIT: The MARMIT model. This model requires the spectra of air-dried soil of the test dataset and a regression model, calibrated on training data from the same soil type. The MARMIT model is not invariant to scaling. As it does not incorporate information from a saturated soil sample, it is expected that a trained model is not or only to a limited extent transferable to other soil types.
- NRAL: The first methodology proposed in this work. Our methodology requires the spectra of an air-dried soil and saturated soil sample (endmembers) and their ground truth SMC for each test dataset. The relationship between normalized relative arc length and the SMC is assumed to be linear.
- SM_S: The second methodology proposed in this work. The relationship between normalized relative arc length and the SMC can be trained on data from the same soil type as the test data, or from one particular soil type. In that case, the spectrometer dataset of Pure Goethite of the Imec-Visionlab_2021 dataset is applied as training data because its SMC has a wide range, varying between 0-101 (g/g × 100).

The quantitative comparisons are provided by the root mean squared error (RMSE), i.e. the error between the estimated SMC ($\hat{\theta}$) and the ground truth SMC (θ):

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\theta}_i - \theta_i\right)^2} \times 100 \qquad (16)$$

where N is the number of test spectra.

We have performed three different types of experiments that required different applications of the datasets. In the first experiment, no variability is introduced. For this, MARMIT and the proposed method were trained and tested on the same data (i.e., data from the same soil type and same illumination and viewing angle). Philpot_2014 and Imec-Visionlab_2021 datasets are suitable for this experiment because they contain sufficient data to be able to train the models.
In the second experiment, we validated the proposed approach for its invariance to a) illumination and viewing angle, and to b) sensor type. As the Liu_2002 dataset contains data acquired at various illumination and viewing angles and our dataset contains data acquired by two different sensor types, these have been applied in this experiment.
In the third experiment, we validated our approach for its near invariance to soil type. For this reason, the model is trained on one particular soil, i.e., Goethite from our dataset, and is tested on all datasets used in this study. We like to clarify that, in this experiment, the regression parameters obtained from the Goethite dataset were fixed and not further calibrated for new soil types.

A. Experiment 1: no variability in illumination and viewing angle and sensor type

In the first experiment, the methods are applied without introducing any variability. For this, all methods are applied on each soil type separately. NRAL, NSMI and SM do not need any training. SM and the proposed methods (NRAL and SM_S) require air-dried soil and saturated soil spectral reflectances and ground truth SMC, while MARMIT requires only air-dried soil data. As MARMIT and SM_S require sufficient training data, we have only applied them on the Philpot_2014 and the Imec-Visionlab_2021 datasets. On these datasets, half of the data was used for training and the other half for testing, and 10 independent runs with different training data were performed. Results are shown in Table III.

The outcomes of the experiments can be summarized as follows:

- The performance of NSMI is poorest. The low performance of NSMI suggests that a single regression model is not valid for every soil type. The performance of NSMI could be improved by fine-tuning the regression parameters (a and b) for each dataset separately.
- SM performed much better than NSMI. This is because the model makes use of information from the the air-dried and saturated soil endmembers. The drawback of this model is that it only utilizes spectral information from the band at 2210 nm.
- MARMIT performs similarly as SM. It does not require the saturated endmember, but it is trained on half of the ground truth.
- NRAL accurately estimated SMC, because it makes use of information from the the air-dried and saturated soil endmembers. The real advantage of NRAL is its invariance to illumination and viewing angle and acquisition conditions, as will be demonstrated in the following experiments.
- SM_S accurately estimated SMC, because it makes use of both air-dried soil and saturated soil endmembers and half of the ground truth. In comparison to NRAL, the performance of SM_S is much better for the Philpot_2014 and Imec-Visionlab_2021 datasets. This suggests that SM_S is suitable for soil-specific calibration. The real advantage of SM_S is its invariance to illumination and viewing angle and acquisition conditions and transferability, as will be demonstrated in the next experiments.

B. Experiment 2: variability in illumination and viewing angle and sensor type

In the next experiment, the invariance of the proposed methods NRAL and SM_S with respect to illumination and viewing angle, and sensor type is validated. For this, both the Liu_2002 and the Imec-Visionlab_2021 datasets are applied. In the former dataset, the illumination and zenith angles vary. To include variability, MARMIT and the proposed method (SM_S) are trained using one particular viewing configuration (illumination angle: 15^0 and zenith angle: 0^0), and tested on all other configurations. Due to lack of sufficient training data, the models are trained on all soil types of that configuration simultaneously. Moreover, the endmembers of air-dried soil (for MARMIT, SM, NRAL, and SM_S) and saturation soil (for SM, NRAL, and SM_S) of that particular configuration are used for all test data. As NSMI is invariant to scaling, its results will be similar for the different configurations, and need not to be shown again. In the Imec-Visionlab_2021 dataset, data from two different sensors, a hyperspectral camera and a spectrometer are available. MARMIT and SM_S are trained with the camera data, and tested on the spectrometer data. In this case, each mineral is trained separately, again 50 % of the camera data is applied for training, and 10 independent runs are performed. Endmembers are taken from the spectrometer data.

Results are shown in Table IV. In Fig. 8, a scatterplot displays the estimated versus ground truth SMC of the four methods on the Liu_2002 dataset. From the results, it is clear that the SM model is not able to capture the introduced variability at all. This is due to the fact that the SM model was developed by assuming that the saturated and dry endmembers are captured under the same conditions as those of the unsaturated spectra. The result of MARMIT is worse than in experiment 1, while the proposed approaches obtained similar results (see Fig. 8 (a) and (b)). From Fig. 8 (c), it can be observed that the MARMIT model shows a serious underestimation for high SMC and overestimation for low SMC. The former is caused by the use of the sigmoid function

TABLE III: The results of different SMC estimation techniques in terms of RMSE (%). For MARMIT and SM_S on both Philpot_2014 and Imec-Visionlab_2021, 50% of the samples have been used for testing (10 runs, mean and standard deviations are shown).

Dataset	MARMIT	NSMI	SM	NRAL (Ours)	SM_S (Ours)
Liu_2002	-	17.17	15.98	4.41	-
Lobell_2002	-	30.04	7.04	5.74	-
Bablet_2016	-	12.25	7.32	6.12	-
Lesaignoux_2008	-	23.04	5.17	3.94	-
Philpot_2014	6.62±2.52	15.58	4.21	3.40	**0.73±0.05**
Imec-Visionlab_2021	15.32±5.14	22.71	17.78	6.27	**3.56±0.50**

TABLE IV: The results of SMC estimation in terms of RMSE (%), on data including variability in illumination and viewing angle and acquisition conditions. On the Imec-Visionlab_2021 dataset, 50% of the samples have been used for training (10 runs, mean and standard deviations are shown).

Dataset	MARMIT	SM	NRAL (Ours)	SM_S (Ours)
Liu_2002	12.88	144.51	3.59	**3.56**
Imec-Visionlab_2021	20.37±2.05	17.78	6.27	**4.08±0.36**

as a regression model, restricting the soil moisture content to be below 40 (g/g $\times$ 100). The latter is caused by the fact that the MARMIT model cannot tackle spectral variability caused by differences in the illumination and viewing angle. The small under-and overestimations in our methods (see Figure 8 (a) and (b)) are caused by the fact that saturated and dry samples from one particular configuration are chosen. This way, some of the saturated samples from other configurations have higher optical path length, and their SMC is restricted to the saturated SMC of the chosen configuration. However, the SMC of saturated samples with lower optical path length than the chosen configuration is underestimated. The inverse argumentation holds for the low SMC samples.

C. Experiment 3: variability in soil type

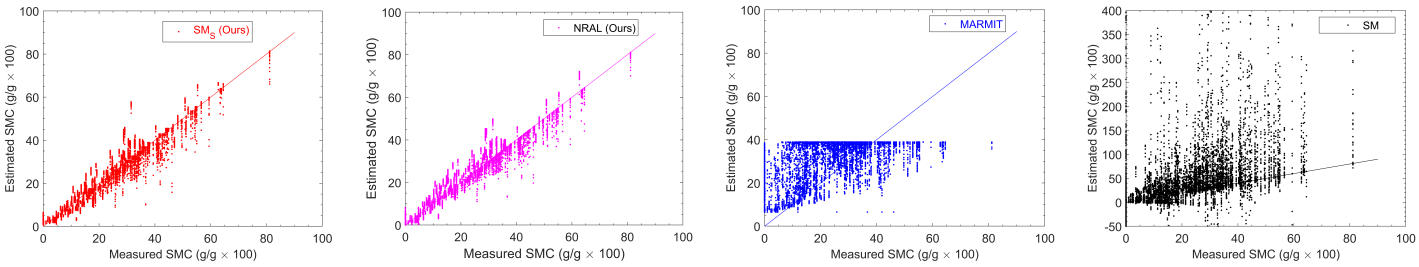
As we already demonstrated in Figure 7, NRAL of moist soils is a measure of soil moisture that is nearly invariant to soil type. To experimentally validate SM_S, in the final experiment, training samples are acquired from one soil type and testing is performed on other soil types. The training is performed solely on the spectrometer data of the Goethite mineral of the Imec-Visionlab_2021 dataset. Results on all datasets are shown in Table V.

Results on the Philpot_2014 and Imec-Visionlab_2021 datasets show that MARMIT performs worse when compared to experiments 1 and 2. This is because the proxy estimated by the the MARMIT model is soil dependent. As a consequence, when trained on one soil type, the models are not transferable to other soil types. SM performed as good as the proposed method for four out of six datasets. However, its performance dropped dramatically on the Liu_2002 (92 soil samples) and the Imec-Visionlab_2021 dataset. It is interesting to observe that NRAL outperformed both MARMIT and SM for all six datasets. This demonstrates that the normalized relative arc length is a nearly invariant measure of soil moisture. SM_S also produced consistent results (RMSE between 2-7 %) in all of the experiments and is the best performer for four out of six datasets.

VI. Discussion

From the experiments, the following general conclusions can be drawn:

- NSMI is an easy method that only requires spectral reflectance at two specific wavelengths. It is based on a regression model, trained on a large number of soil types. It is scale-invariant, and thus invariant to most variability in illumination and viewing angle. In general, NSMI does not perform well in estimating SMC from moist soil samples. Its major drawbacks are that it only makes use of limited spectral information, and that it does not account for the structure of water in soils. To improve its performance, the parameters of the model (a and b) should be fine-tuned for each dataset separately, requiring ground truth training samples, which makes it impractical in real world scenarios.
- The SM model only uses very limited spectral information, but does not need any training. It requires the spectral reflectance of both an air-dried and saturated soil sample, partially accounting for the distribution of the water in the soil. It is not invariant to scaling. SM performed very well for four out of six datasets without variability in illumination and viewing angle but completely failed in the case of variability in illumination and viewing angle. For the SM model to perform well, the spectra of the air-dry, moist and saturated sample should be acquired with the same illumination/viewing configuration. This makes the use of the SM model impractical in real world scenarios. When applied outdoors, reflectance spectra are inevitably impaired by variations caused by surface topography and illumination conditions. Variations in illumination conditions cause random scaling effects in spectral reflectances. This makes single-band models less robust for predicting soil moisture content because they cannot remove the scaling factor from the spectral reflectance. Because the random scaling factor is the same for all bands, most soil moisture indices (e.g.,



(a) (b) (c) (d)

Fig. 8: Ground truth SMC vs. estimated SMC of moist soils (Liu_2002 dataset). (a) SM_S (Ours), (b) NRAL (Ours), (c) MARMIT, and (d) SM.

TABLE V: The results of SMC estimation on all datasets in terms of RMSE (%). Training is performed on Goethite from the spectrometer data of the Imec-Visionlab_2021 dataset. In the latter dataset, RMSE is computed only for Iron, Lime, Silica, and White Sand.

Dataset	MARMIT	SM	NRAL (Ours)	SM_S (Ours)
Liu_2002	48.14	15.98	4.41	**4.36**
Lobell_2002	44.68	7.04	**5.74**	7.52
Bablet_2016	53.80	7.32	6.12	**4.04**
Lesaignoux_2008	42.01	5.17	**3.94**	5.59
Philpot_2014	28.23	4.21	3.40	**2.63**
Imec-Visionlab_2021	51.52	17.78	6.27	**5.02**

NSMI) utilize spectral reflectance from at least two bands to cancel out the scaling factor.

- MARMIT requires training and the use of the air-dried soil spectral reflectance. It is not invariant to scaling. As MARMIT is learning a regression between the parameter $L \times \epsilon$ and SMC, this relation is soil-dependent. When trained on each soil separately, MARMIT performs equally well as SM, but this approach is impractical in real world situations. When tested on other soil types, the trained model fails. The performance may be improved by training specific models for similar soils. In [3], training was performed separately on distinct classes of soils, based on their textural, mineralogical, and spectral properties. To significantly improve the performance of the MARMIT model, one should partially account for the the distribution of the water in the soil, by utilizing the relation between $L \times \epsilon$ and SMC of a saturated soil sample.
- The proposed approach SM_S was found to be the most consistent and performed well for almost all soil datasets used in this study. This demonstrates the potential of the proposed methodology, in particular, in cross-sensor situations and in case of spectral variability caused by variations in acquisition and illumination conditions. By utilizing the spectrum of the saturated soil and its ground truth SMC the estimated proxy becomes nearly invariant to the soil type.
- NRAL performed as well as SM_S, without having to perform any training. The results suggest that SM_S is more suitable for soil-specific calibration when the accuracy requirement is very high (RMSE $<$ 3.5 %) while NRAL can be applied when the error in moisture content estimation can be in the range of 3-6 %.
- In the third group of experiments, it was demonstrated that the trained model is transferable from one soil type to any of the applied datasets from the literature. The only mineral for which the proposed approach does not seem to work was pure SiO_2. Although silica is present in almost all soils, even in high percentages, in its pure form, it is quite challenging to produce a homogeneous mixture of water and silica, because of its hydrophobic nature. In this scenario, the optically active thickness of the layer of water can vary with the different incident angles of the light. Mixing it with even a very low percentage of other hydrophilic minerals is sufficient to reduce this effect.
- In outdoor environment, the strong absorption bands of molecular water around 1400 nm and 1900 nm are masked by atmospheric absorption ([6], [8]). To demonstrate that the method still works in field conditions, we removed these absorption regions and only utilized reflectance values between 1118-1300 nm and 1501-1654 nm on the Liu_2002 dataset in experiment 2. The drop in performance was less than 0.3%.
- Even though our method was found to be robust to external variability, the problem will be challenging in case of spatial variations of the physical properties and chemical composition of the soil. To demonstrate its applicability in highly heterogeneous scenarios, three similar soil samples (PrairieA, PrairieB, and PrairieC) from the Lesaignoux_2008 dataset were processed. For both PrairieA and PrairieC, dry and saturated spectra of PrairieB were utilized for estimating the normalied relative arc length. The algorithm was trained on Goethite and tested on moist soil samples of PrairieA and PrairieC.

The performance of SM_S for estimating SMC decreased by less than 1% for PrairieA and less than 3% for PrairieC.

- In comparison to microwave radiation-based methods, most of the reflectance-based methods, including our method, are limited to bare soil. A major challenge would be to remove the effects of vegetation on the spectral reflectance.

All methods were developed in MATLAB and ran on an Intel Core i7-8700K CPU, 3.20 GHz machine with six cores. The runtime of NSMI, SM, and NRAL was below 1s. MARMIT and SM_S require training, which took approximately 1s for both MARMIT and SM_S, when trained on the Goethite data. Testing of both models is done within one second.

VII. Conclusion

In this work, we developed a normalized relative arclength (NRAL) as a proxy for soil moisture content, which is invariant to illumination and viewing angle, and sensor type. Moreover, the proxy is made less dependent on soil type, by utilizing the spectral reflectance of the saturated soil and its ground truth SMC. Next to NRAL,a supervised approach (SM_S) was proposed for soil moisture content estimation. The proposed approaches were validated and compared to a number of methods from the literature on a large number of soil data, including data with variable illumination and viewing angles and in cross-sensor situations. The results suggest that the proxy NRAL can be used as an estimate of SMC. For the supervised approach (SM_S), the trained models were demonstrated to be transferable from one soil type to another soil type. The only drawback of this work is the requirement of saturated soil information. In future work, we will attempt to remove this drawback. Although the proposed methods are claimed to be less dependent on soil type, some initial measurements showed that their performance dropped in the case of hydrophobic minerals. This should be investigated in more detail.

Acknowledgement

The research presented in this paper is funded by BELSPO (Belgian Science Policy Office) in the frame of the STEREO III programme – project GEOMIX (SR/06/357) and by the Research Foundation-Flanders - project G031921N. The authors would like to thank the company Engie-Laborelec for the use of their hyperspectral SWIR camera, and Prof. R. Samson for the use of the spectrometer.

References

[1] R.R. Deshmukh S. Ansari, "Estimation of soil moisture content: a review," *Int. J. Theoret. Appl. Mech.*, vol. 12, no. 3, pp. 571–577, 2017.

[2] Sonia I. Seneviratne, Thierry Corti, Edouard L. Davin, Martin Hirschi, Eric B. Jaeger, Irene Lehner, Boris Orlowsky, and Adriaan J. Teuling, "Investigating soil moisture–climate interactions in a changing climate: A review," *Earth-Science Reviews*, vol. 99, no. 3, pp. 125–161, 2010.

[3] A. Bablet, P.V.H. Vu, S. Jacquemoud, F. Viallefont-Robinet, S. Fabre, X. Briottet, M. Sadeghi, M.L. Whiting, F. Baret, and J. Tian, "Marmit: A multilayer radiative transfer model of soil reflectance to estimate surface soil moisture content in the solar domain (400–2500nm)," *Remote Sensing of Environment*, vol. 217, pp. 1–17, 2018.

[4] Ebrahim Babaeian, Morteza Sadeghi, Scott B. Jones, Carsten Montzka, Harry Vereecken, and Markus Tuller, "Ground, proximal, and satellite remote sensing of soil moisture," *Reviews of Geophysics*, vol. 57, no. 2, pp. 530–616, 2019.

[5] Zhao-Liang Li, Pei Leng, Chenghu Zhou, Kun-Shan Chen, Fang-Cheng Zhou, and Guo-Fei Shang, "Soil moisture retrieval from remote sensing measurements: Current knowledge and directions for the future," *Earth-Science Reviews*, vol. 218, pp. 103673, 2021.

[6] Weidong Liu, F. Baret, Xingfa Gu, Bing Zhang, Qingxi Tong, and Lanfen Zheng, "Evaluation of methods for soil surface moisture estimation from reflectance data," *International Journal of Remote Sensing*, vol. 24, no. 10, pp. 2069–2083, 2003.

[7] Liu Weidong, F. Baret, Gu Xingfa, Tong Qingxi, Zheng Lanfen, and Zhang Bing, "Relating soil surface moisture to reflectance," *Remote Sensing of Environment*, vol. 81, no. 2, pp. 238–246, 2002.

[8] D.B. Lobell and G.P. Asner, "Moisture effects on soil reflectance," *Soil Science Society of America Journal*, vol. 66, no. 3, pp. 722–727, 2002.

[9] Pariva Dobriyal, Ashi Qureshi, Ruchi Badola, and Syed Ainul Hussain, "A review of the methods available for estimating soil moisture and its implications for water resource management," *Journal of Hydrology*, vol. 458-459, pp. 110–117, 2012.

[10] S.-N. Haubrock, S. Chabrillat, C. Lemmnitz, and H. Kaufmann, "Surface soil moisture quantification models from reflectance data under field conditions," *International Journal of Remote Sensing*, vol. 29, no. 1, pp. 3–29, 2008.

[11] Edwin T. Engman, "Applications of microwave remote sensing of soil moisture for water resources and agriculture," *Remote Sensing of Environment*, vol. 35, no. 2, pp. 213–226, 1991.

[12] Jia Tian and William D. Philpot, "Relationship between surface soil water content, evaporation rate, and water absorption band depths in swir reflectance spectra," *Remote Sensing of Environment*, vol. 169, pp. 280–289, 2015.

[13] Attila Nagy, Péter Riczu, Bernadett Gálya, and János Tamás, "Spectral estimation of soil water content in visible and near infra-red range," *Eurasian Journal of Soil Science*, pp. 163 – 171, 2014.

[14] D. Zhang and G. Zhou, "Estimation of soil moisture from optical and thermal remote sensing: A review," *Sensors (Basel, Switzerland)*, vol. 16, 2016.

[15] R. M. Hoffer and C. J. Johannsen, "Ecological potential in spectral signatures analysis," *Remote Sensing in Ecology*, pp. 1–16, 1969.

[16] Morteza Sadeghi, Scott B. Jones, and William D. Philpot, "A linear physically-based model for remote sensing of soil moisture using short wave infrared bands," *Remote Sensing of Environment*, vol. 164, pp. 66–76, 2015.

[17] Anders Ångström, "The albedo of various surfaces of ground," *Geografiska Annaler*, vol. 7, pp. 323–342, 1925.

[18] Walter G. Planet, "Some comments on reflectance measurements of wet soils," *Remote Sensing of Environment*, vol. 1, no. 2, pp. 127–129, 1970.

[19] S. A. Bowers and R. J. Hanks, "Reflection of radiant energy from soils," *Soil Science*, vol. 100, pp. 130–138, 1965.

[20] D. L. NEEMA, AJAY SHAH, and A. N. PATEL, "A statistical optical model for light reflection and penetration through sand," *International Journal of Remote Sensing*, vol. 8, no. 8, pp. 1209–1217, 1987.

[21] D.G. Levitt, J.R. Simpson, and A.R. Huete, "Estimates of surface soil water content using linear combinations of spectral wavebands," *Theoretical and Applied Climatology*, vol. 42, no. 4, pp. 245–252, 1990, cited By 12.

[22] Zhongling Gao, Xingang Xu, Jihua Wang, Hao Yang, Wenjiang Huang, and Huihui Feng, "A method of estimating soil moisture based on the linear decomposition of mixture pixels," *Mathematical and Computer Modelling*, vol. 58, no. 3, pp. 606–613, 2013, Computer and Computing Technologies in Agriculture 2011 and Computer and Computing Technologies in Agriculture 2012.

[23] Jia Tian, Jibo Yue, William D. Philpot, Xinyu Dong, and Qingjiu Tian, "Soil moisture content estimate with drying process segmentation using shortwave infrared bands," *Remote Sensing of Environment*, vol. 263, pp. 112552, 2021.

[24] Audrey Lesaignoux, Sophie Fabre, and Xavier Briottet, "Influence of soil moisture content on spectral reflectance of bare soils in the 0.4–14 µm domain," *International Journal of Remote Sensing*, vol. 34, no. 7, pp. 2268–2285, 2013.

[25] Zhe Yin, Tingwu Lei, Qinghong Yan, Zhanpeng Chen, and Yuequn Dong, "A near-infrared reflectance sensor for soil surface moisture measurement," *Computers and Electronics in Agriculture*, vol. 99, pp. 101–107, 2013.

[26] Michael L Whiting, Lin Li, and Susan L Ustin, "Predicting water content using gaussian model on soil spectra," *Remote Sensing of Environment*, vol. 89, no. 4, pp. 535–552, 2004.
[27] J. Peng, H. Shen, and S.W. He, "Soil moisture retrieving using hyperspectral data with the application of wavelet analysis," *Environ Earth Sci*, vol. 69, pp. 279–288, 2013.
[28] A. M. Mouazen, R. Karoui, J. De Baerdemaeker, and H. Ramon, "Characterization of soil water content using measured visible and near infrared spectra," *Soil Science Society of America Journal*, vol. 70, no. 4, pp. 1295–1302, 2006.
[29] H. Bach and W. Mauser, "Modelling and model verification of the spectral reflectance of soils under varying moisture conditions," in *Proceedings of IGARSS '94 - 1994 IEEE International Geoscience and Remote Sensing Symposium*, 1994, vol. 4, pp. 2354–2356.
[30] William Philpot, "Spectral reflectance of wetted soils," *Proceedings of ASD and IEEE GRS*, vol. 2, pp. 1–12, 2010.
[31] P. Von Kubelka, "Ein beitrag zur optik der farbanstriche," *Zeitschrift fur technische Physik*, vol. 12, pp. 593–601, 1931.
[32] H. Bach, *Die Bestimmung hydrologischer und landwirtschaftlicher Oberflächenparameter aus hyperspektralen Fernerkundungsdaten, Münchener Geographische Abhandlungen*, 1995.
[33] Kent F. Palmer and Dudley Williams, "Optical properties of water in the near infrared∗," *J. Opt. Soc. Am.*, vol. 64, no. 8, pp. 1107–1110, Aug 1974.
[34] Bikram Koirala, Zohreh Zahiri, Alfredo Lamberti, and Paul Scheunders, "Robust supervised method for nonlinear spectral unmixing accounting for endmember variability," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–15, 2020.
[35] C.E. Rasmussen and C.K.I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, New York, 2006.

Bikram Koirala (Member, IEEE) received the M.S. degree in geomatics engineering from the University of Stuttgart, Stuttgart, Germany, in 2016, and the Ph.D. degree in development of advanced hyperspectral unmixing methods from the University of Antwerp, Antwerp, Belgium, in 2021.

In 2017, he joined the Vision Lab, Department of Physics, University of Antwerp, as a Ph.D. Researcher, where he is currently a Post-Doctoral Researcher. His research interests include machine learning and hyperspectral image processing.

Zohreh Zahiri received the bachelor degree in Civil Engineering from Shahid Ashrafi Esfahani University, Isfahan, Iran, and the M.S. degree in Architectural Conservation from the Isfahan University of Art in 2009 and in 2013 respectively. She received her Ph.D. degree in "Characterization of facade materials using close-range hyperspectral data" from University College Dublin (UCD), Department of Civil Engineering, Dublin, Ireland. She is currently a postdoctoral researcher at Vision Lab, Department of Physics, University of Antwerp. Her research interests are the application of hyperspectral imaging in civil engineering and related areas as well as data analysis methods.

Paul Scheunders (M'98) received the B.S. degree and the Ph.D. degree in physics, with work in the field of statistical mechanics, from the University of Antwerp, Antwerp, Belgium, in 1983 and 1990, respectively. In 1991, he became a research associate with the Vision Lab, Department of Physics, University of Antwerp, where he is currently a full professor. His current research interest includes remote sensing and hyperspectral image processing. He has published over 200 papers in international journals and proceedings in the field of image processing, pattern recognition, and remote sensing. Paul Scheunders is Associate Editor of the IEEE Transactions on Geoscience and Remote Sensing and has served as a program committee member in numerous international conferences. He is a senior member of the IEEE Geoscience and Remote Sensing Society.